贝叶斯多组比较——渐近测量不变性

张沥今 [1]；宋琼雅 [1]；潘俊豪 [1*]

[1] 中山大学心理学系，广州，510006

摘　要　测量工具的测量不变性是在潜变量框架下进行多组比较的前提条件。贝叶斯渐近测量不变性方法基于贝叶斯思想的优良特性，通过为参数的跨组差异提供合适的先验分布，放宽了传统的多组验证性因子分析方法对跨组差异的严格限制。同时避免了传统方法容易导致模型拟合过差、修正过程繁琐及一类错误率上升等问题，具有极高的应用价值。文章总结并介绍了渐近测量不变性方法的原理及优势，同时通过实例展示了该方法在 M*plus* 软件中的具体分析过程。

关键词　贝叶斯方法；多组验证性因子分析；渐近测量不变性

1　引言

在心理学、教育学等社会科学领域中，群体间差异的检验一直受到广泛的关注，如探究不同年龄的儿童在认知能力上的差异(欧阳湘子，田伟，辛涛，詹沛达，2016)，不同文化背景下人群的行为、态度的差异(Shorey, Allan, Cohen, Fite, Stuart, & Temple, 2018)等等。而在这类研究中，研究者所关注的变量常常是抽象的构念(Construct)，即不能直接被观测的、而需要用量表条目等可以直接观测的变量来反映的潜变量(Latent Variables)，如态度、价值观、能力等，其中用于反映潜变量的量表条目等指标则被称为可观察变量(Observed Variables)。

在潜变量的框架下进行多组比较时，测量工具的测量不变性(Measurement Invariance, MI)是进行组间差异比较的前提条件。测量不变性是指测量工具所测量的构念不会因为施测群体的改变而改变，即来自不同群体的参与者对于相同条目的反应范式是相同的（Mellenbergh, 1989; van de Schoot, Kluytmans, Tummers, Lugtig, Hox, & Muthén, 2013）。以潜变量框架为基础的结构方程模型为测量不变性的检验提供了切实可行的方法。结构方程建模通常包含测量模型和结构模型两部分，其中，测量模型主要反映可观察变量与潜变量之间的关系，通常为一个验证性因子分析(Confirmatory Factor Analysis, CFA)模型。测量不变性的检验可以通过结构方程建模的测量模型部分来实现。

---

* 国家自然科学基金项目(31871128)；

作者备注：张沥今、宋琼雅为共同第一作者

通讯作者：潘俊豪，E-mail: panjunh@mail.sysu.edu.cn

目前已经发展出了多种检验测量不变性的结构方程建模方法，包括多组验证性因子分析模型(例如 Raju, Laffitte, & Byrne, 2002; Stark, Cherryshenko, & Drasgow, 2006)，多层验证性因子分析模型(Jak, Oort, & Dolan, 2013)，多层因子混合模型(Kim, Cao, Wang, & Nguyen, 2017)，对齐法(Asparouhov & Muthén, 2014)及拓展的对齐法(Marsh et al., 2018)等等。其中，多组验证性因子分析（多组 CFA，Multiple Group CFA）是最为常用的方法(Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002 等)，而本文将针对传统的多组 CFA 方法和贝叶斯多组 CFA 方法进行深入地讨论。

1.1 多组验证性因子分析

多组 CFA 建模需要对每个组别分别建立一个验证性因子分析模型，并在此基础上比较组间反应范式的差异(Kim et al., 2017; Rutkowski & Syetina, 2014)。在多组 CFA 中检验测量不变性需要依次施加测量模型参数跨组相等的限制，按顺序包括(Schmitt & Kuljanin, 2008)：（1）形态不变性(Configural Invariance)模型，要求各组具有相同的因子结构，即量表条目在每组中均被负载到相同的潜变量上；（2）载荷不变性(Metric Invariance)，又称弱测量不变性模型，指不同群组的因子载荷跨组相等，即潜变量每变化一个单位在可观察变量上表现出的变化是跨组相等的；（3）截距不变性(Scalar Invariance)又称强测量不变性模型，指条目截距跨组相等，即不同群组的测量具有相同的参照点。只有在满足单位和参照点都跨组相等时（载荷和截距不变性），因子均值的跨组比较才是有意义的，否则可能导致有偏的估计结果(Schmitt & Kuljanin, 2008)；（4）误差方差恒定(Error-variance Invariance)，又称严格的测量不变性模型，指误差方差跨组等值，这意味着可观察变量分数变异的跨组差异完全反映了因子分变异的跨组差异(Rutkowski & Svetina, 2014)。

多组 CFA 方法在每施加一种测量不变性限制后，都需要通过模型拟合比较来检验施加限制后模型拟合是否显著变差，在满足了当前不变性限制的基础上才能够进行下一种跨组不变性的检验。针对这种嵌套模型的拟合比较可以采用卡方差异检验(Chi-square difference test)，它根据模型的卡方值和自由度的变化判断施加限制后模型拟合是否显著变差。但由于卡方差异检验容易受到样本量的影响，一些研究者提出了更加有效的模型拟合比较标准，例如ΔCFI$\leqslant$0.01，ΔRMSEA$\leqslant$0.03 等(Little & Card, 2013; Rutkowski & Svetina, 2014)。

多组 CFA 方法思路简洁且易于掌握，得到了广泛的应用。在多组 CFA 模型满足跨组截距不变性的基础上，研究者还可以进一步进行结构不变性的检验，包括对因子均值(Factor Mean)、因子方差/协方差(Factor Variance/Covariance)的跨组比较(Little & Card, 2013)。目前

流行的结构方程建模软件也都支持多组 CFA 建模，便于研究者应用该方法。

1.2 多组验证性因子分析的局限

尽管多组 CFA 方法存在着诸多优势且被广泛应用，但是在实证研究中，多组 CFA 方法对于跨组不变性的限制往往过于严苛，要求载荷、截距严格的跨组相等，这使得添加限制条件后的模型更容易被拒绝(Asparouhov & Muthén, 2014; Kim et al., 2017)。Marsh 等人(2018)指出在实证研究中截距不变性模型几乎不会被满足，且这种现象在组数较多的情况下更为严重，极大地影响了研究的后续分析(Rutkowski & Svetina, 2014)。

在实证研究中，当载荷或截距的跨组不变性不被满足时，研究者通常可以采用事后模型修正(Post-hoc Modification)方法，结合理论和修正指数(Modification Index)的建议释放一些违背跨组不变性的参数以改善模型拟合情况，即建立部分测量不变性(Partial Measurement Invariance)模型。但是建立部分测量不变性模型的方法依然存在着一些局限：（1）基于修正指数的事后模型修正方法每释放一个参数都需要进行模型拟合比较，当需要修正的参数过多时，修正过程会极为繁琐；（2）通过修正指数发现的违背测量不变性的参数可能是由于抽样变异性导致的(Asparouhov & Muthén, 2014)，同时，参数的释放还会受到研究者主观性的影响，这增加了一类错误率，降低了结论的可重复性；（3）在这种部分测量不变性模型的基础上进行组间比较，容易导致对跨组因子均值差异的有偏估计(Marsh et al., 2018)。

多组 CFA 方法的上述局限为应用研究带来了较大的阻碍，且这种问题在组数较多或违背不变性的参数较多时更为严重。因此，方法学家一直在探索和开发具有更高准确性与适用性的多组比较方法。Asparouhov 和 Muthén(2013)提出的贝叶斯渐近测量不变性(approximate Measurement Invariance, approximate MI)方法，较好地解决了因传统方法对跨组限制过于严格而产生的问题。遗憾的是，由于其发展时间较短，应用研究者对于这种建模思路并不熟悉、对其相关软件操作也不够了解。本文将对该方法进行详尽地介绍与总结，阐释该方法的原理与优势。同时以实例分析演示如何在实证研究中应用渐近测量不变性方法，并与传统多组 CFA 方法进行对比。希望本文能够为应用研究者提供新的多组建模思路，解决采用传统方法建模时难以克服的问题。

2 渐近测量不变性

2.1 方法原理

在结构方程建模中，除了传统的频率学派估计方法(如，极大似然估计)，贝叶斯估计方法近年来也愈加流行(王孟成，邓俏文，毕向阳，2017；张沥今，陆嘉琦，魏夏琰，潘俊豪，2018；van de Schoot et al., 2017)。相比于传统方法在小样本、非正态数据和复杂模型等情况下容易出现模型无法识别、参数估计有偏等问题，贝叶斯结构方程建模对模型或数据的极端情况更为宽容(李锡钦，2011)。

从统计学思想的角度来看，传统的频率学方法将未知参数视为常数，而贝叶斯方法则将其视为随机变量，结合样本数据和先验信息得到未知参数的后验分布(王孟成等，2017)。因此贝叶斯方法能够灵活地纳入先验信息（如，前人研究或预实验结果），而有效的先验信息可以带来更准确的参数估计(Yuan & MacKinnon, 2009)。研究者对先验信息准确性的把握体现在先验信息的强度中，即先验分布的方差大小中。先验分布的方差越小，对未知参数后验分布的影响越大。例如，结构方程模型中被限制为 0 的参数就可以被视为提供了均值为 0，方差为 0 的先验分布。

鉴于贝叶斯方法在结构方程建模中表现出的优良特性，越来越多的方法学家和应用研究者致力于发展和应用贝叶斯结构方程模型(van de Schoot et al., 2017)，进而衍生出了一系列新的建模思路。例如，传统方法在进行结构方程建模时，需要根据理论假设测量模型中不存在交叉载荷或测量误差相关，即严格限制这些参数为 0，但是这种严格的限制在实际应用中往往很难得到满足(Muthén & Asparouhov, 2012)。基于贝叶斯方法能够结合先验信息的特性，Muthén 和 Asparouhov(2012)创造性地指出：如果为交叉载荷和测量误差相关提供均值为 0、方差极小的先验分布，就可以允许这些参数在 0 附近波动，放宽对模型的严格限制。而放宽模型限制后发现的显著不为 0 的交叉载荷或测量误差相关，可以被视为由抽样变异性导致的、不需要理论解释的冗余参数(nuisance parameters)。这种放宽模型限制的思想自提出后就得到了深入的研究与推广，研究者们基于这种思想进一步提出了诸多优化方法(如 Lu, Chow, & Loken, 2016；Pan, Ip, & Dubé, 2017 等)。Muthén 和 Asparouhov(2013)更进一步将其应用到了多组结构方程建模中，发展出渐近测量不变性方法。

在传统的多组 CFA 建模中，我们需要对形态不变性模型依次施加跨组载荷相等、跨组截距相等的严格限制。而正如前文所述，这种限制在实证研究中几乎不会被满足(Marsh et al., 2018)。尽管此时可以选择释放部分违背测量不变性的参数，但是这种部分测量不变性的方法又存在着诸多局限。而在渐近测量不变性分析中，通过对载荷和截距的跨组差异提供均值为 0、方差极小的先验分布，可以放宽传统方法对其的严格限制，允许载荷和截距存在微小的跨组差异。Muthén 和 Asparouhov(2013)指出，渐近测量不变性方法本质上是认为数据存

在着诸多微小、随机的跨组差异，这些差异可能是互为相反方向的、可以相互抵消的，因此不影响对模型结构不变性的检验。渐近测量不变性方法可以在允许存在微小跨组差异的同时，实现对结构方程模型参数的准确估计。

基于放宽模型限制的思想和贝叶斯方法的优良特性，渐近测量不变性方法相比于传统方法有着诸多优势：（1）避免了传统方法限制过于严格而容易导致模型拟合过差的问题；（2）在一次估计中就可以放宽对所有参数的限制，而不需要进行多次事后模型修正。避免了修正时重复利用同一批数据可能导致的更高的一类错误率、有偏的参数估计等问题(Draper, 1995)；（3）可以作为一种检验违背测量不变性的参数的工具(Muthén & Asparouhov, 2013)：渐近测量不变性方法能够一次性发现所有违背不变性的参数(non-invariant parameters)；（4）通过提供不同的小方差先验分布可以检验数据违背不变性的程度(Cieciuch, Davidov, Algesheimer, & Schmidt, 2017)等。

模拟研究也显示，在载荷或截距参数存在着较小的跨组差异时，贝叶斯渐近测量不变性方法在模型拟合和参数估计中表现都很好(Kim et al., 2017; van de Schoot et al., 2013)，且渐近测量不变性方法基于贝叶斯结构方程模型的框架，更具灵活性，模型拟合评估也更为有效(Kim et al., 2017; Schmitt & Kuljanin, 2008)。

2.2 分析步骤

渐近测量不变性方法自提出以来，就得到了广泛的发展和应用(De Bondt & Van Petegem, 2015; van de Schoot et al., 2013 等)。但由于其发展时间较短，具体的分析流程和评价标准尚未规范，本文基于渐近测量不变性方法的一些应用研究和模拟研究结果(Fong & Ho, 2014; Kim et al., 2017; van de Schoot et al., 2013 等)，总结出了较为完整、全面的分析步骤：

2.2.1 形态不变性

和传统方法一样的是，在建立渐近测量不变性模型时，首先需要建立基线模型(Baseline Model)，即形态不变性模型。在模型拟合良好的前提下，进一步施加载荷和截距跨组相等的限制。模型拟合良好指后验预测 p 值(Posterior Predictive p-value, PPp 值)在 0.5 左右，且后验预测区间(Posterior Predictive p-value Interval)包括 0。其中，PPp 值在 0.5 左右的判断标准较为主观，模拟研究显示通常可以采用 0.1 的截断值(Meghan & Zhang, 2018; Muthén & Asparouhov, 2012)，即 PPp 值大于 0.1 且后验预测区间包括 0 时可以认为模型拟合良好。

此外，在形态不变性的检验中，还可以根据 Muthén 和 Asparouhov(2012)放宽模型限制的思路，同时放宽对于模型交叉载荷或测量误差相关的严格限制(Fong & Ho, 2014)，即对这

些参数提供均值为 0、方差极小的先验分布。

2.2.2 载荷和截距不变性

在满足形态不变性的基础上，传统方法需要按顺序依次施加载荷和截距跨组不变性的限制。这主要是由于传统方法在出现模型拟合显著变差，即模型违背跨组不变性限制时，需要通过事后模型修正方法逐个释放违背不变性的参数。在这种情况下，如果同时施加载荷和截距跨组相等的限制，会导致引起模型拟合变差的潜在因素过多，影响对违背不变性的参数的寻找(Little & Card, 2013)。

渐近测量不变性方法则可以通过对载荷和截距的跨组差异提供均值为 0、方差极小的先验分布，在一次估计中放宽对所有载荷和截距跨组相等的严格限制。因此在渐近测量不变性分析中，可以同时施加载荷和截距跨组不变性的限制，即对载荷和截距的跨组差异同时提供均值为 0，方差极小 (建议采用 0.01 的方差，即在数据被标准化处理的前提下，允许跨组差异有 95%的变异落在(-0.2, 0.2)之间) 的正态先验分布(Muthén & Asparouhov, 2013)。如果此时模型拟合较差，则表示数据不能满足不变性假设，而如果模型拟合较好还需要进一步进行敏感性分析。

2.2.3 敏感性分析

van de Schoot 等人(2013)及 Kim 等人(2017)均建议在采用渐近测量不变性分析时，需要进行敏感性分析(Sensitivity Analysis; Greenland, 2001)以避免先验方差的选取受到研究者主观性的影响。

具体来说，研究者需要先确定预设先验(predetermined prior)方差，建议采用 0.01 (Kim et al., 2017)。在施加跨组不变性限制、对载荷和截距的跨组差异提供 *N*(0, 0.01)的先验分布后可以得到模型和数据的拟合程度，如 M*plus* 中提供的 PP*p* 值、偏差信息准则(Deviance Information Criterion, DIC)和贝叶斯信息准则(Bayesian Information Criterion, BIC)。这一步骤即为上文载荷和截距不变性步骤。

在 PP*p* 值显示模型拟合良好的前提下，对跨组差异提供方差更大的先验分布，通常为 *N*(0,0.05) (Kim et al., 2017; van de Schoot et al., 2013)。通过模型拟合比较含有两种先验分布的模型：如果结果显示先验分布为 *N*(0,0.01)的模型拟合更好、或两个模型没有显著差异，则可以认为模型满足跨组载荷和截距不变性；若 *N*(0,0.01)的先验分布下模型拟合显著更差，则说明对模型施加跨组不变的限制是错误的，导致了模型拟合明显变差，即模型不满足跨组不变性(Kim et al., 2017)。

在进行模型拟合比较时，Kim 等人(2017)建议采用 DIC 指标。因为 BIC 指标总是会选择先验分布方差为 0.05 的模型，这可能是由于方差 0.05 允许存在的跨组差异相对更大，通常会发现更多显著违背不变性的参数，而 DIC 则会惩罚过于复杂的模型。且 DIC 更加符合贝叶斯偏差(Bayesian Deviance)的概念(Kaplan & Depaoli, 2012)，因而更多地被用于贝叶斯估计中。在分析中，当两个模型 DIC 差异的绝对值大于 7 时，则有足够的证据认为 DIC 更小的模型优于另一个模型，否则可以认为两个模型没有明显差异(李锡钦，2011；Spiegelhalter, Best, Carlin, & Van Der Linde, 2002)。

在满足载荷和截距跨组不变性的基础上，研究者可以进行后续因子均值的比较，探究多群组在同一构念上表现的差异。尽管渐近测量不变性方法目前已经可以满足大部分实证研究的需求，但还没有合适的先验分布能够放宽对测量误差方差/协方差跨组不变的严格限制，因此渐近测量不变性方法还无法支持组间方差/协方差不变性的检验。希望随着贝叶斯方法的蓬勃发展，未来可以有更好的方法解决这一局限。

2.3 部分渐近测量不变性

贝叶斯渐近测量不变性方法假设载荷和截距参数中存在着诸多微小的组间差异，在这种数据条件下该方法可以实现准确的参数估计。但是如果数据中一些载荷或截距参数存在着较大的组间差异，*N*(0, 0.01)的小方差先验可能会严重压缩这些参数真实的组间差异，进而造成对跨组均值差异的有偏估计(Muthén & Asparouhov, 2013)。因此研究者们建议在发现显著违背测量不变性的参数后，可以采用部分渐近测量不变性(Partial approximate Measurement Invariance)方法对这些参数进行自由估计，以实现更准确的跨组因子均值比较(Muthén & Asparouhov, 2013; van de Schoot et al., 2013)。

在传统的多组 CFA 中，部分测量不变性方法是指当严格的载荷或截距不变性模型不被满足时，研究者可以找出违背了跨组不变性的载荷或截距参数，对其进行自由估计，即建立部分测量不变性模型。在这种部分测量不变性模型的基础上也可以进一步检验结构不变性(Schmitt & Kuljanin, 2008)，当违背测量不变性的参数数目相对较少时，部分测量不变性对结构模型参数估计的准确性影响较小(Little & Card, 2013)。

类似的，在贝叶斯渐近测量不变性方法中，通过对跨组差异提供 *N*(0,0.01)的先验分布，可以检测到所有跨组差异显著不为 0，即违背测量不变性的参数。通过对这些参数去除先前提供的 *N*(0,0.01)先验分布，采用 M*plus* 默认的无信息先验分布，即对其进行自由估计，就可以建立部分渐近测量不变性模型(Muthén & Asparouhov, 2013)。相比于传统方法，贝叶斯部

分渐近测量不变性方法能够在一次估计中检测到所有违背跨组不变性的参数，极大地降低了事后模型修正方法的工作量，且避免了研究者主观地选择需要修正的参数。模拟研究也显示，贝叶斯部分渐近测量不变性方法可以获得更准确的参数估计结果(van de Schoot et al., 2013)。

但是贝叶斯部分渐近测量不变性模型依然存在着与传统部分测量不变性模型相同的两个主要问题：（1）需要释放的参数可能是由于抽样变异性导致的；（2）释放违背跨组不变性的参数会对结构模型参数的估计产生怎样的影响尚不确定(Schmitt & Kuljanin, 2008)。因此，研究者也需要结合理论假设，建立更符合理论意义的部分渐近测量不变性模型。同时也希望未来有更多的模拟研究能够探究部分测量不变性对结构模型参数估计的影响，为应用研究者提供更准确的建议。

3　示例：结构方程模型中贝叶斯渐近测量不变性分析

为了检验不同性别在社会支持量表上反应的差异，本研究采用传统的多组 CFA 和贝叶斯渐近测量不变性两种方法进行建模，展示贝叶斯渐近测量不变性方法的具体分析步骤，并与多组 CFA 方法进行对比。研究采用 M*plus*8.0 软件进行建模，贝叶斯渐近测量不变性方法对应的 M*plus* 代码详见附录。数据来源于 353 名本科大学生，平均年龄为 19.55 岁(*SD* = 1.45，range:17-24) ，其中男生 89 名。

研究采用叶悦妹和戴晓阳(2008)编制的大学生社会支持评定量表，量表共 17 题，包括主观支持、客观支持和支持利用度三个维度，每个维度分别有 5 题、6 题和 6 题，采用 Likert 5 点评分。题项如“大多数同学都很关心我”、“面对两难的选择时，我会主动向他人寻求帮助”等。本研究中该问卷内部一致性系数为 0.928，三个子维度的内部一致性系数分别为 0.892、0.818 和 0.881，问卷信度良好。

以往研究表明，内部一致性系数需要假设观察变量间的测量误差相互独立，否则可能产生有偏的信度系数估计结果(Raykov, 2001)。而组合信度(Composite Reliability)允许测量误差之间存在一定的相关, 且允许各条目在因子上的载荷不同，更适用于潜变量(徐万里, 2008)。根据 Fornell 和 Larcker(1981)的建议组合信度为 0.6 以上即证明量表信度较好。本研究中量表三个子维度的组合信度分别为 0.894、0.819 和 0.882，即量表子维度的组合信度也表现良好。

3. 1　多组验证性因子分析

根据偏度与峰度分析结果，各条目的偏度在-1.107 到-1.43 之间，峰度在-0.705 到 1.534

之间。依照偏度和峰度检验标准，一般认为偏度绝对值大于 2 和峰度绝对值大于 7 为非正态分布(West, Finch, & Curran, 1995)。因此，本研究中条目均基本符合正态分布，且由于数据中不含缺失值，可以采用极大似然法(Maximum Likelihood，ML)进行模型估计。在多组 CFA 建模中，依次建立跨组形态不变性模型、载荷不变性模型和截距不变性模型。根据 Little 和 Card(2013)的建议，在模型比较中如果 ΔCFI 的绝对值小于 0.01 可以认为两个模型的拟合无明显差异。

在形态不变性建模中，发现两组中分别存在一对残差相关，对其进行自由估计后建立形态不变性模型作为基线模型，模型拟合良好(表 1)。进一步建立载荷不变性模型发现模型不满足跨组载荷相等(CFI = 0.917, |ΔCFI| = 0.04)，采用事后模型修正方法放宽对两个条目[1]载荷跨组相等的限制，即对其进行自由估计，建立部分载荷不变性模型，模型拟合和形态不变性模型之间没有明显差异(CFI = 0.920, |ΔCFI| = 0.01)。在部分载荷不变性模型的基础上检验发现模型不满足截距不变性(CFI = 0.909, |ΔCFI| = 0.11)。在释放了对三个条目[2]截距跨组相等的限制后，建立部分截距不变性模型，模型拟合和部分载荷不变性模型没有明显差异(CFI = 0.919, |ΔCFI| = 0.01)。

在部分截距不变性模型的基础上，比较两组因子均值差异发现，女生在主观支持、客观支持和支持利用度三个维度上得分均显著高于男生(标准化均值差异：*dif* = 0.389, p = 0.002; *dif* = 0.566, p < 0.001; *dif* =0.410, p = 0.003)。

表 1 多组验证性因子分析模型拟合及模型比较

模型	χ^2	*df*	CFI	TLI	SRMR	RMSEA	90%CI of RMSEA	ΔCFI	Pass?
形态不变性	494.817	230	0.921	0.907	0.064	0.081	[0.071,0.091]		
载荷不变性	521.450	244	0.917	0.908	0.075	0.080	[0.071,0.090]	-0.04	No
部分载荷不变性	512.170	242	0.920	0.910	0.072	0.080	[0.070,0.089]	-0.01	Yes
截距不变性	560.015	256	0.909	0.904	0.083	0.082	[0.073,0.091]	-0.11	No
部分截距不变性	526.292	253	0.919	0.913	0.076	0.078	[0.069,0.088]	-0.01	Yes

[1] 放宽的条目为支持利用度维度下的两题，题项为“当我遇到麻烦时，我通常会主动寻求别人的帮助”和“当有烦恼时，我会主动向同学、朋友倾诉”。

[2] 放宽的条目包括客观支持维度下的两题，题项为“在困难的时候，我可以依靠家人或亲友”、“我经常能得到家人、亲友的照顾和支持”；及支持利用度维度下的一题，题项为“当有烦恼时，我会主动向家人、亲友倾诉”。

3.2 贝叶斯渐近测量不变性

根据Muthén和Asparouhov(2012)的建议，对误差项矩阵提供*IW*(*I*, *df*)的逆Wishart分布，其中*I*为单位矩阵，*df*为*p*+6，其中*p*为可观察变量数目（本例中*df* = 23），即允许每个残差协方差参数有95%的变异落在(-0.2, 0.2)之间，构建形态不变性模型，发现模型拟合良好(表2)。

分别对载荷和截距的跨组差异提供 *N*(0,0.01)的先验分布，结果显示，载荷和截距不变性模型拟合良好。进一步进行敏感性分析，对载荷和截距的跨组差异提供 *N*(0,0.05)的先验分布，结果显示两种不同的先验分布下模型拟合无明显差异(|ΔDIC| = 2.252) (表2)，即数据满足载荷和截距的跨组不变性。此外，在放宽了对跨组差异的严格限制后，“DIFFERENCE OUTPUT”输出显示存在一个截距参数[3]违背了跨组不变性，对该参数进行自由估计后建立部分渐近测量不变性模型(表2)。

表2 贝叶斯渐近测量不变性分析--模型拟合及模型比较

模型	PP*p*	95%C.I.[1]	BIC	DIC	ΔDIC[2]
形态不变性	0.624	[-86.681, 62.040]	14723.613	13131.384	
载荷和截距不变性(先验方差0.01)	0.664	[-90.916, 60.739]	14766.800	13113.782	
载荷和截距不变性(先验方差0.05)	0.674	[-90.851, 57.596]	14757.127	13111.530	-2.252
部分测量不变性	0.681	[-93.554, 58.890]	14764.332	13112.122	

注：[1] 95%C.I.为观察的卡方值与重复抽样获得的卡方值间差异的置信区间。

[2] ΔDIC为敏感性分析检验先验方差为0.01和0.05的两种模型的DIC差异。

在部分渐近测量不变性模型的基础上，比较两组因子均值同样发现女生在主观支持、客观支持和支持利用度三个维度上得分均显著高于男生，(标准化均值差异：*dif* = 0.394, *p* = 0.004; *dif* = 0.443, *p* < 0.001; *dif* =0.559, *p* < 0.001)。与Muthén和Asparouhov(2012)及van de Schoot等人(2013)的研究结果一致的是，渐近测量不变性方法对因子均值差异的估计结果与传统方法相差不大。

---

[3] 违背不变性的截距参数对应的条目属于客观支持维度下，题项为“在困难的时候，我可以依靠家人或亲友”。

3.3 方法比较

从上述分析中可以发现，传统的多组 CFA 方法对模型限制过于严格，且模型比较过程和修正过程都更为繁琐。此外，多组 CFA 方法中现有的比较标准仍然存在着较多争议(Little & Card, 2013)，容易受到研究者主观性的影响。而当载荷不变性或截距不变性模型不被满足时，事后模型修正方法显示有更多的违背不变性的载荷和截距参数，模型修正还可能存在研究者对修正参数进行主观选择的偏差。此外，未来还需要更多的模拟研究来探索部分测量不变性方法对后续因子均值差异的估计产生的影响。

而贝叶斯渐近测量不变性方法操作简单，通过放宽对组间差异的严格限制也避免了传统方法容易导致模型拟合过差的问题，敏感性分析也可以帮助研究者避免先验信息主观性的影响，得到更可靠的结果。

4 总结与展望

基于贝叶斯渐近测量不变性方法的优良特性，该方法也越来越受到应用研究者的欢迎。例如，Cieciuch 等人(2017)将贝叶斯渐近测量不变性方法应用于欧洲社会调查中，采用施瓦茨价值观量表(21-item Portrait Value Questionnaire，PVQ-21)测量个人世界观并比较其跨国差异。研究者采用了 15 个国家中 6 次测量的数据，通过比较传统多组 CFA 方法和贝叶斯渐近测量不变性方法，发现贝叶斯渐近测量不变性方法能够使更多的国家满足测量不变性，利于后续的因子均值比较。该方法也同样被应用于焦虑、抑郁因子的性别不变性检验(Fong & Ho, 2014)，伴侣间亲密关系态度差异的检验(Chiorri, Day, & Malmberg, 2014)等应用研究中。希望本文的介绍能帮助研究者更快地掌握这一方法，解决实际数据分析中遇到的一些障碍。

但是作为一种建模工具，贝叶斯渐近测量不变性方法不可避免地存在着一定的适用范围和局限：首先，贝叶斯渐近测量不变性方法在组数多时计算耗时更长，效率相对较低；其次，当模型不能满足载荷和截距不变性时应当如何处理？目前还没有较为完整的结论，研究者们提出了诸如部分渐近测量不变性(Muthén & Asparouhov, 2013)、对齐法等诸多解决办法(Marsh et al., 2018)，但这些方法都还存在着许多亟待发展的地方，未来还需要更多的模拟研究和实证研究来填补这些空白。最后，尽管渐近测量不变性方法能够完成对载荷和截距的跨组不变性检验，也可以用于后续的因子均值比较,但是目前还不能检验误差方差/协方差不变性。

除了贝叶斯渐近测量不变性方法，目前也存在着其他检验组间测量不变性的诸多方法，包括：多层验证性因子分析，多层因子混合模型，对齐法及拓展的对齐法等等。这些方法虽

然有其较为独特的优势，但是适用范围相对更较窄，也存在许多局限。例如多层分析方法要求组数较多且参数的组间差异较大，在适用范围上不及贝叶斯渐近测量不变性方法(Kim et al., 2017)；而对齐法及拓展的对齐法只要求模型满足形态不变性，不需要进行后续模型拟合比较，不利于对违背不变性参数的寻找(Kim et al., 2017; Marsh et al., 2018 等)。针对现有不同方法的优劣比较及其适用范围的讨论依然有待探究，在测量不变性的基础上，这种方法间的比较也可以进一步扩展到结构不变性层面，包括进行因子均值，因子方差/协方差的不变性检验。

参考文献

李锡钦. (2011). *结构方程模型: 贝叶斯方法* (蔡敬衡, 潘俊豪, 周影辉译). 北京: 高等教育出版社.

欧阳湘子, 田伟, 辛涛, & 詹沛达. (2016). IRT 框架下追踪数据的测量不变性分析——以 4 至 5 岁儿童认知能力测验为例. *心理科学*(3), 606-613.

王孟成, 邓俏文, 毕向阳. (2017). 潜变量建模的贝叶斯方法. *心理科学进展, 25*(10), 1682–1695.

徐万里. (2008). 结构方程模式在信度检验中的应用. *统计与信息论坛, 23*(7), 9-13.

叶悦妹, 戴晓阳. (2008). 大学生社会支持评定量表的编制. *中国临床心理学杂志. 16(5): 465-468.*

张沥今, 陆嘉琦, 魏夏琰, 潘俊豪.(2018). 贝叶斯结构方程模型及其研究现状. *心理科学进展*, doi:10.12074/201812.00865

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508.

Cieciuch, J., Davidov, E., Algesheimer, René, & Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European social survey. *Sociological Methods & Research*, *47*(4), 665-686.

Chiorri, C., Day, T., & Malmberg, L. E. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Frontiers in Psychology, 5*, 983.

De Bondt, N., & Van Petegem, P. (2015). Psychometric Evaluation of the Overexcitability Questionnaire-Two Applying Bayesian Structural Equation Modeling (BSEM) and Multiple-Group BSEM-Based Alignment with Approximate Measurement Invariance. *Frontiers in Psychology, 6*, 1963.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 45-97.

Fong, T., & Ho, R. (2014). Testing gender invariance of the hospital anxiety and depression scale using the classical approach and bayesian approach. *Quality of Life Research, 23*(5), 1421-1426.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Greenland, S. (2001). Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Analysis, 21*(4), 579-583.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(2), 265-282.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673).

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: a comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 524-544.

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian Factor Analysis as a Variable-Selection Problem: Alternative Priors and Consequences. *Multivariate Behavioral Research, 51*(4), 519-539.

Little, T. D., & Card, N. A. (2013). *Longitudinal structural equation modeling*. The Guilford Press.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B. , Asparouhov, T. , & Muthén B, et al. (2018). What to do when scalar invariance fails: the extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods, 23*(3), 524-545.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(1), 60-72.

Meghan K. Cain & Zhang Z. Y. (2018). Fit for a Bayesian: an evaluation of PP*p* and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4): 1-12.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*(2), 127-143.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313-335.

Muthén, B., & Asparouhov, T. (2013). *BSEM measurement invariance analysis*: Mplus Web Note 17. http://www.statmodel.com/examples/webnotes/webnote17.pdf

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods, 22*(4), 687-704.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-29.

Raykov, T. (2001). Bias of coefficient alpha fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*(1), 69-76.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational & Psychological Measurement, 74*(1), 31-57.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: review of practice and implications. *Human Resource Management Review, 18*(4), 210–222.

Shorey, R. C., Allan, N. P., Cohen, J. R., Fite, P. J., Stuart, G. L., & Temple, J. R. (2018). Testing the Factor Structure and Measurement Invariance of the Conflict in Adolescent Dating Relationship Inventory.

Psychological Assessment. Advance online publication. http://dx.doi.org/10.1037/pas0000678

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

Stark, S. , Chernyshenko, O. S. , & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306.

van de Schoot. R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with scylla and charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology, 4*, 770.

van de Schoot. R., Winter, S. D., Ryan, O., Zondervanzwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: the last 25 years. *Psychological Methods, 22*(2), 217-239.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. (pp. 56-75). Thousand Oaks, CA: Sage Publications, Inc.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14*(4), 301-322.

Bayesian multiple-group analysis: approximate measurement invariance

ZHANG Lijin[1]; SONG Qiongya[1]; PAN Junhao[1]

([1] *Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China*)

Abstract: Measurement invariance (MI) is a pre-requisite for comparison between groups under the framework of latent variables. But in traditional multi-group factor analysis, scalar invariance is almost unachievable in practice. The Bayesian approximate MI proposed by Muthén and Asparouhov (2013) compensates for this limitation to some extent by providing a zero-mean, small-variance prior for the differences in measurement parameters. It allows for small differences between groups and avoids the problem caused by strict restrictions in classical method, such as poor model fitting, awkward model modifications and higher Type I error rate. These strengths make this new approach a more suitable method in the practical research. Throughout the paper we summarized the principles and advantages of the approximate MI approach, and a real data set was analyzed to demonstrate the validity and practical usefulness of this approach using M*plus*.

Key words: Bayesian estimation; multiple-group structural equation modeling; approximate measurement invariance

附录：贝叶斯渐近测量不变性分析：*Mplus* 代码

1. 形态不变性模型

```
TITLE: Configural Model
DATA: FILE = socialsupport.dat;
VARIABLE:
        NAMES = Gender SS1-SS17;
        USEV = SS1-SS17;
        KNOWNCLASS IS c (Gender=1-2); !按性别分为两组，1 指男性，2 指女性
        CLASSES IS c(2);    !组数为 2
ANALYSIS:
        MODEL=CONFIGURAL;    !形态不变性模型，允许跨组载荷、截距自由估计
        TYPE=MIXTURE;
        ESTIMATOR = BAYES;
        PROC = 2;
        BITERATIONS = 100000(10000);
          !最小迭代次数 10000 次，最大迭代次数 100000 次，模型收敛即停止迭代
MODEL:
      %OVERALL%
      F1 BY SS1-SS5*;      !*放宽对第一个观察指标载荷为 1 的限制
      F2 BY SS6-SS11*;
      F3 BY SS12-SS17*;
      F1-F3@1;    !采用固定方差法识别模型


      SS1-SS17(P#_1-P#_17);
          !定义测量误差参数，P#_1 指第#组的第一个测量误差参数(# = 1/2)
      SS1-SS17 WITH SS1-SS17(P#_18-P#_153);    !定义测量误差相关参数
MODEL PRIORS:    !对参数提供先验分布
      DO(1,2)P#_1-P#_17~IW(1,23);
      DO(1,2)P#_18-P#_153~IW(0,23);
          ! DO(1,2)代表循环，指#参数从 1 循环到 2
```

!对两组的误差项矩阵均提供 *IW*(*I*, 23)分布，放宽对测量误差相关的严格限制

```
OUTPUT:   TECH1   TECH8   STDY   SVALUES;
PLOT: TYPE= PLOT2;
```

2. 载荷和截距不变性模型

```
TITLE: Scalar Model
DATA: FILE = socialsupport.dat
VARIABLE:
      NAMES = Gender SS1-SS17;
      USEV = SS1-SS17;
      KNOWNCLASS IS c (Gender=1-2);
      CLASSES IS c(2);
ANALYSIS:
      MODEL=ALLFREE;   !对两组的所有参数进行自由估计
      TYPE=MIXTURE;
      ESTIMATOR = BAYES;
      PROC = 2;
      BITERATIONS = 100000(10000);
MODEL:
    %OVERALL%
    F1 BY SS1-SS5*(lam#_1-lam#_5);
    F2 BY SS6-SS11*(lam#_6-lam#_11);
    F3 BY SS12-SS17*(lam#_12-lam#_17);
    [SS1-SS17](nu#_1-nu#_17);
    SS1-SS17(P#_1-P#_17);
    SS1-SS17 WITH SS1-SS17 (P#_18-P#_153);

    %c#1%    !设定第一组为参照组，设定参照组因子方差为 1，均值为 0
    F1-F3@1;
    [F1-F3@0];
```

```
%c#2%    !对第二组中因子均值和方差进行自由估计
F1-F3;
[F1-F3];

MODEL PRIORS:
    DO(1,2)P#_1-P#_17~IW(1,23);
    DO(1,2)P#_18-P#_153~IW(0,23);
    DO(1,17)DIFF(lam1_#-lam2_#)~N(0,0.01);
    DO(1,17)DIFF(nu1_#-nu2_#)~N(0,0.01);
        !对两组载荷、截距差异提供N(0, 0.01)的先验分布
        !DIFF(lam1_#-lam2_#) 指第一组和第二组中的第#个载荷参数的差异值
OUTPUT:  TECH1  TECH8 STDY  SVALUES;
PLOT: TYPE= PLOT2;
```

3. 部分渐近测量不变性模型

```
MODEL PRIORS:
    DO(1,2)P#_1-P#_17~IW(1,23);
    DO(1,2)P#_18-P#_153~IW(0,23);
    DO(1,17)DIFF(lam1_#-lam2_#)~N(0,0.01);
    DO(1,7)DIFF(nu1_#-nu2_#)~N(0,0.01);
    DO(9,17)DIFF(nu1_#-nu2_#)~N(0,0.01);
        !Mplus 会提供无信息先验分布对第 8 个截距参数进行自由估计
        !模型其余部分与载荷和截距不变性模型相同
```